

Validación estadística de la presencia en plantas de quimiotipos caracterizados por la concentración de componentes volátiles obtenida mediante GC-MS

Ana Cristina Soria*, Joaquín Esteban*, Ramón Morales**, Pedro J. Martín-Álvarez*** y Jesús Sanz*

Resumen: Soria, A. C.; Esteban, J.; Morales, R.; Martín-Álvarez, P. J. & Sanz, J. 2008. Validación estadística de la presencia en plantas de quimiotipos caracterizados por la concentración de componentes volátiles obtenida mediante GC-MS. *Bot. Complut.* 32: 225-236.

Se denomina “quimiotipo” a un grupo de individuos de una especie que se distingue de forma significativa del resto por su composición química. Mientras que la existencia de polimorfismo químico es bien conocida para diversas especies, el significado estadístico de la diferencia entre sus composiciones no ha recibido suficiente atención. Se propone, para la asignación de un nivel de significado a la existencia de un quimiotipo caracterizado por un compuesto, un método basado en la formación de agrupaciones mediante el procedimiento de k-medias, el cálculo de la distancia de Mahalanobis D_{mah} como estimación de la separación entre grupos, y la comparación de este valor con valores D_{mah} calculados a partir de datos simulados con una distribución normal. El método se aplica a los datos de concentración obtenidos por GC-MS de compuestos volátiles de muestras de *Thymus zygis* subsp. *zygis*, *Thymus zygis* subsp. *sylvestris* y *Lavandula luisieri*.

Palabras clave: quimiotipos, componentes volátiles, k-medias, distancia de Mahalanobis, *Thymus zygis* subsp. *zygis*, *Thymus zygis* subsp. *sylvestris*, *Lavandula luisieri*.

Summary: Soria, A. C.; Esteban, J.; Morales, R.; Martín-Álvarez, P. J. & Sanz, J. 2008. Statistical validation of the presence in plants of chemotypes characterized from their volatile component concentration obtained by GC-MS. *Bot. Complut.* 32: 225-236.

The name “chemotypes” indicates a group of individuals of a species which can be significantly distinguished from the rest from their chemical composition. Whereas the chemical polymorphism of several species is well known, the statistical significance of their differences in composition has been scarcely studied. The application of the k-means algorithm as a clustering technique, of the Mahalanobis distance D_{mah} as a measure of group separation and its comparison with D_{mah} values calculated from simulated data having a normal distribution, is proposed in order to assign a significance level to a single compound chemotype. The method is applied to volatile concentration data from samples of *Thymus zygis* subsp. *zygis*, *Thymus zygis* subsp. *sylvestris* and *Lavandula luisieri*.

Keywords: chemotypes, volatile components, k-means, Mahalanobis distance, *Thymus zygis* subsp. *zygis*, *Thymus zygis* subsp. *sylvestris*, *Lavandula luisieri*.

INTRODUCCIÓN

La determinación de los componentes químicos presentes en las plantas ha permitido poner de manifiesto que, para ciertas especies, individuos indistinguibles desde un punto de vista botánico y que crecen en similares condiciones atmosféricas y edáficas presentan una composición química diferente. A partir de este hecho, se ha propuesto el término “quimiotipo” para calificar a

un grupo infraespecífico de individuos que se caracterizan y distinguen de modo significativo de los demás miembros de su especie por la presencia o concentración de uno o varios compuestos, así como el de “quimiotaxonomía” para los estudios que utilizan la diversidad química de las plantas para complementar los resultados obtenidos del conocimiento de su diversidad botánica. El estudio de la existencia de quimiotipos presenta un interés aplicado cuando la utilidad de una especie

* Instituto de Química Orgánica General, (CSIC), Juan de la Cierva, 3, 28006 Madrid. acsoria@ifi.csic.es.

** Real Jardín Botánico, (CSIC), Plaza de Murillo, 2, 28014 Madrid. *** Instituto de Fermentaciones Industriales (CSIC), Juan de la Cierva, 3, 28006 Madrid

Recibido: 5 marzo 2008. Aceptado 21 abril 2008.

se basa en las propiedades de algunos de sus compuestos, cuyas concentraciones en los distintos individuos podrían variar de forma notable. Su interés científico se deriva fundamentalmente de la posibilidad de emplear la composición química como indicadora de la existencia de una diversidad genética que se pone de manifiesto no en diferencias morfológicas sino en características metabólicas.

La presencia de un componente (compuesto marcador), ausente en otros individuos de la especie, puede emplearse a veces para definir un quimiotipo. Más habitual es considerar como característico un compuesto cuando alcanza un determinado nivel de concentración, generalmente relativo. También puede basarse la diferenciación entre quimiotipos en perfiles de composición que incluyen varios componentes o sus relaciones. En cualquiera de estos casos, y de acuerdo con la definición previamente mencionada, la propuesta de la existencia de quimiotipos debería contemplar varios criterios. En primer lugar, las variaciones de composición observadas entre individuos no deben ser asignables a factores propios de la toma de muestra, de su preparación y de su análisis. Debe contemplarse la posibilidad de que sean causadas por factores edáficos, estacionales o climáticos. Es conveniente disponer de un número de datos analíticos para distintos individuos de la especie, que permitan estimar, para varios compuestos, la dispersión de los valores de concentración obtenidos. Por último, la diferencia entre esos valores para las muestras de distintos quimiotipos debe ser significativa en comparación con dicha dispersión. El empleo de los compuestos volátiles de las plantas con fines quimiotaxonómicos permite utilizar en la mayor parte de los casos un elevado número de componentes, entre los que es más probable la presencia de compuestos marcadores o característicos. Por otro lado, el acoplamiento cromatografía de gases-espectrometría de masas (GC-MS) es especialmente adecuado para proporcionar de una forma rápida la necesaria información cualitativa y cuantitativa sobre los componentes volátiles de una planta. Para facilitar el procesado y evaluación del elevado volumen de información analítica generada, es necesario recurrir al empleo de técnicas de análisis estadístico.

En las técnicas de análisis multivariante no supervisadas se busca la presentación de información numérica compleja en una forma más fácilmente interpretable. Las más utilizadas en el campo de la quimiotaxonomía han sido el análisis de componentes principales (PCA) y

el análisis de conglomerados (CA). Ambas permiten postular la presencia de agrupaciones entre las muestras estudiadas, aunque no proporcionan información cuantitativa sobre su significado. Las técnicas supervisadas, como el análisis discriminante, que suministran esta información no pueden utilizarse a priori con este propósito, al requerir una información previa sobre la clasificación incompatible con la definición de quimiotipo. Por este motivo, las propuestas acerca de la presencia de quimiotipos en una especie suelen partir de unas agrupaciones definidas de forma subjetiva, sin contemplar su significado estadístico. La mayor parte de los programas estadísticos utilizados en el estudio de agrupaciones usan como punto de partida su existencia y no consideran la distribución de los datos. Tampoco se ha profundizado en el significado estadístico de los quimiotipos que se proponen. De hecho, el reconocimiento de la presencia de quimiotipos debería basarse en que las diferencias observadas entre ellos no puedan adscribirse a factores aleatorios y sean, por tanto, significativas. En el caso de un quimiotipo definido por un compuesto marcador, la distribución de dicho compuesto debería ser bimodal, mientras que un caso de referencia tendría una distribución gaussiana normal de tipo unimodal, no compatible con la existencia de quimiotipos. Para la comparación entre ambos casos debería emplearse un parámetro estadístico, relevante para la distribución en grupos. El valor de este parámetro para los datos experimentales se compararía con la distribución de referencia, calculando la probabilidad de que procediese de una distribución de este tipo por causas aleatorias (Lawson & Jurs 1990).

En este trabajo, presentamos un método para la estimación objetiva de la probabilidad de que las variaciones en concentración de un compuesto para distintas plantas individuales de una especie correspondan a la existencia de una población (quimiotipo) para cuyos individuos la concentración de dicho compuesto difiere de forma significativa de la de los restantes miembros de la especie. Para ello, utilizamos como hipótesis nula la suposición de que la tendencia de los datos experimentales a formar dos grupos de distinta concentración es la que cabría esperar de una distribución normal. Como método objetivo para la obtención de los dos grupos, tanto en los datos experimentales como en los de referencia, hemos elegido el método de agrupación de *k*-medias. Como medida de la separación entre grupos, se ha utilizado la distancia de Mahalanobis, cuyo valor para los datos experimentales se

ha comparado con los calculados para distribuciones normales de referencia, obtenidas mediante simulación. Si de acuerdo con esta comparación se rechaza la hipótesis, la agrupación puesta de manifiesto en los datos experimentales no es debida a causas aleatorias y podemos defender la existencia de un quimiotipo. El método se ha aplicado a la detección de quimiotipos en datos de concentración procedentes del análisis por GC-MS de componentes volátiles en plantas de *Thymus zygis* Löffl. ex L. subsp. *zygis*, *Thymus zygis* subsp. *sylvestris* (Hoffm. & Link) Brot. ex Cout. y *Lavandula luisieri* (Rozeira) Rivas Mart. La distribución de concentraciones presenta en cada una de estas tres especies diferentes tipos de pautas. Como sistema de introducción de muestra se ha empleado la desorción térmica directa acoplada en línea a cromatografía de gases con detector de masas (DTD-GC-MS) para reducir el tiempo de análisis y permitir su aplicación a un elevado número de muestras. Los géneros *Thymus* y *Lavandula* se caracterizan por su riqueza en componentes volátiles y por la presencia de quimiotipos en sus especies, habiendo sido estudiados tanto por grupos españoles (Velasco-Negueruela *et al.* 1984, 1985, García-Martín & García-Vallejo 1979, 1983, García-Vallejo 1992, García-Vallejo *et al.* 1992, Morales Valverde 1986, Sáez 1995) como extranjeros (Stahl-Biskup 1991, Salgueiro 1994, Michet *et al.* 2008).

MATERIALES Y MÉTODOS

Muestras de plantas: Todas las muestras fueron recogidas en la Península Ibérica, bajo control botánico para confirmar su pertenencia a la especie y en su etapa de floración. En la Comunidad de Madrid se muestrearon 48 individuos de *Thymus zygis* subsp. *zygis* y 50 de *Thymus zygis* subsp. *sylvestris*: una pequeña cantidad de hojas y flores se tomó de cada individuo muestreado. 50 muestras de *Lavandula luisieri* se recogieron en Toledo y Sevilla, y flores y hojas se analizaron por separado. Las muestras se dejaron secar a temperatura ambiente antes de su análisis. **Desorción térmica directa:** La fracción volátil se obtuvo mediante un sistema ATD 400 (Perkin-Elmer, Norwalk, CT, USA). Entre 2-20 mg de planta seca se introdujeron en un tubo de PTFE, que se colocó en un cartucho de desorción de acero inoxidable. La desorción se llevó a cabo a 180°C durante 15 min, en un flujo de helio de 20 mL min⁻¹. Los volátiles se criofocaron en una trampa de Tenax a -30°C. Para la inyección, la trampa se calentó hasta 320°C a 40°C s⁻¹, y los volátiles desorbidos se transfirieron a la columna cromatográfica mediante un tubo calentado de sílice fundida. El procedimiento se detalla en Esteban *et al.* (1993). **GC-MS:** El ATD 400 se conectó a un cromatógrafo de gases GC 8000 (Fisons, Milan, Italy) acoplado a un detector de masas MD 800 (Fisons, Manchester, UK). Las columnas, de metilpolisiloxano (20-30 m x 0.25 mm d.i. x 0.25 mm espesor de

fase), se programaron entre 60 y 180°C a 3°C min⁻¹ y luego hasta 250°C a 8°C min⁻¹. Se usó helio a ~ 1 mL min⁻¹ como gas portador. Los espectros de masas se registraron mediante impacto electrónico a 70 eV en el intervalo *m/z* 38-350. El procesamiento de los datos se llevó a cabo mediante el programa MassLab version 1.4 (Finnigan, Manchester, UK). **Análisis cualitativo y cuantitativo:** A partir del perfil de corriente total de iones (TIC), los compuestos volátiles detectados se caracterizaron mediante sus espectros de masas, que se compararon con los de las bibliotecas Wiley (McLafferty & Stauffe 1989) y NIST (NIST/EPA/NIH 2002). Las posibles identificaciones se confirmaron mediante la retención cromatográfica, usando índices de retención lineales para los patrones disponibles. Los valores de concentración relativos se calcularon directamente a partir de las áreas de pico TIC en relación al área total. Como posibles indicadores de quimiotipos se seleccionaron aquellos compuestos cuya concentración relativa superaba el 5% en al menos una muestra. También se han seleccionado algunos compuestos que no alcanzaban dicho nivel cuando por su estructura se consideró de interés el estudiar su comportamiento.

Metodología estadística: *Análisis multivariante exploratorio de los datos:* Para tratar de representar de forma gráfica la complejidad de los datos obtenidos, se ha utilizado el análisis de componentes principales (PCA), partiendo de la matriz de covarianza de los datos. Los dos primeros componentes obtenidos se han utilizado para la representación gráfica. Para el *Thymus zygis* subsp. *zygis* se ha empleado, con fines comparativos, el análisis de agrupaciones (cluster análisis, CA), utilizando como distancia entre muestras la euclídea y como método de unión el de Ward.

Datos experimentales y simulados: Para cada una de las plantas estudiadas se empleó como matriz de datos la formada por los valores de concentración relativa (% frente al total de volátiles) de cada uno de los compuestos volátiles (variables) en las muestras estudiadas. Como referencia frente a estos datos experimentales se simuló matrices de datos compuestas por 1000 variables, en cada una de las cuales distintos números de muestras presentaban una distribución normal por variable. *Propuesta de la existencia de un posible quimiotipo a partir de datos de concentración:* Para la división en grupos se ha elegido el algoritmo de agrupación denominado k-medias, (MacQueen 1967, Hartigan 1975, Hartigan & Wong 1979) ya que no requiere la toma de decisiones subjetivas. El algoritmo se ha aplicado a los datos experimentales empleando sucesivamente como variable la concentración relativa de cada compuesto. El objetivo es la separación de las muestras en dos grupos, correspondientes a un posible quimiotipo y al resto de individuos de la especie, de forma que la dispersión dentro de cada uno de ellos sea mínima, mientras que la diferencia entre ambos sea máxima. Para ello, el algoritmo mueve de forma iterativa las muestras de uno a otro grupo, buscando minimizar la varianza dentro de los grupos. El proceso puede repetirse para asegurar que el máximo alcanzado es global. Se obtiene así una clasificación objetiva en dos grupos; de los miembros de cada grupo se calculan sus medias y las diferencias entre éstas.

Existencia de una separación entre grupos significativa: Como medida de la separación entre grupos, se ha utilizado la distancia de Mahalanobis D_{mah} (Mahalanobis 1936, Hartigan 1975):

$$D_{mah} = (\bar{x}_1 - \bar{x}_2) / \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}}$$

siendo las varianzas s_j^2 ,

$$s_j^2 = \sum_{i=1}^{n_j} (x_{i,j} - \bar{x}_j)^2 / (n_j - 1)$$

Su valor para los datos experimentales se ha comparado con los calculados para las matrices de referencia formadas por distribuciones normales. Se ha elegido como hipótesis a rechazar el que los datos experimentales sigan una distribución normal unimodal, siendo las diferencias con la referencia debidas a causas aleatorias.

Corrección de Bonferroni por número de compuestos: El relativamente alto número de compuestos analizado para una muestra es una ventaja, pero introduce un factor aleatorio adicional en los datos, por lo que puede ser conveniente el emplear una estimación más conservadora. Con este objetivo, hemos utilizado la corrección de Bonferroni (Shafer 1995). En ella, se supone que la probabilidad de alcanzar por azar un nivel de significación determinado para una variable aumenta con el número de variables disponibles, por lo que se recomienda el empleo de $\alpha' = \alpha / n$, donde α es el nivel de significación buscado y n el número de variables. En nuestro caso, n es el número de compuestos y α' es el nivel de significación de que, para un valor dado D_{mah} , los grupos encontrados para un compuesto cualquiera provengan por causas aleatorias de una distribución normal. Para los tres conjuntos de datos utilizados, se ha tomado $\alpha = 0.0001$ (nivel de confianza de un 99.99%). Para valores de D_{mah} superiores al valor crítico, se rechaza la hipótesis de una distribución normal, justificándose el significado de los grupos obtenidos. En las tablas correspondientes a los conjuntos de datos estudiados, se señalan con un asterisco los compuestos con D_{mah} superiores al valor crítico y con dos asteriscos los superiores al doble de ese valor.

Programas de cálculo estadístico: El programa STATISTICA (2005) se utilizó para los análisis estadísticos multivariantes llevados a cabo (análisis jerárquico de agrupaciones, análisis de componentes principales) y para la comprobación de resultados. Para la formación de agrupaciones mediante el procedimiento de k-medias y para las simulaciones de distribuciones y cálculo de valores D_{mah} se escribieron programas específicos en VisualBasic 4.0 para Windows.

RESULTADOS

Distribución de los valores de D_{mah} , distancia de Mahalanobis, para los dos grupos obtenidos mediante el procedimiento de las k-medias para una distribución normal

En la Tabla 1 se indican el valor medio y la desviación típica obtenidos para la distancia de Mahalanobis, D_{mah} , entre los dos grupos obtenidos mediante el procedimiento de las k-medias. Las distribuciones se han simulado para valores de muestras comprendidos entre

Tabla 1

Diferencia entre grupos obtenidos mediante k-medias a partir de distribuciones normales simuladas. Valor medio y desviación típica de las distancias de Mahalanobis obtenidas para distintos valores del número de muestras

<i>Distancia de Mahalanobis</i>		
<i>Nº de Muestras (n)</i>	<i>Valor medio</i>	<i>Desv. Típica</i>
10	3.302	0.831
12	3.179	0.722
14	3.122	0.651
16	3.072	0.587
18	3.009	0.528
20	2.993	0.486
25	2.921	0.420
30	2.882	0.386
35	2.856	0.350
40	2.845	0.314
45	2.815	0.306
50	2.799	0.282
60	2.780	0.257
70	2.784	0.244
80	2.769	0.218
90	2.751	0.202
100	2.731	0.191
120	2.737	0.187
140	2.719	0.162
160	2.722	0.155
200	2.706	0.130
250	2.702	0.119
300	2.696	0.113
400	2.692	0.090
500	2.695	0.086
600	2.684	0.078
800	2.679	0.069
999	2.680	0.061

10 y 1000, empleando 1000 variables en cada caso. Se observa que los valores D_{mah} disminuyen al aumentar n , de forma cada vez menos marcada para n elevado. Para el uso de los valores de esta tabla debe por tanto tenerse en cuenta el número de muestras. Si este número es muy próximo a alguno de los de la primera columna de la tabla, pueden utilizarse los valores correspondientes de media y desviación típica. Si no es así, estos valores pueden calcularse por interpolación en función del número de muestras n empleando las ecuaciones: $2.6842 \cdot \exp(2.095/n)$ para el valor medio, y $3.335 \cdot n^{(-0.6238)}$ para la desviación típica.

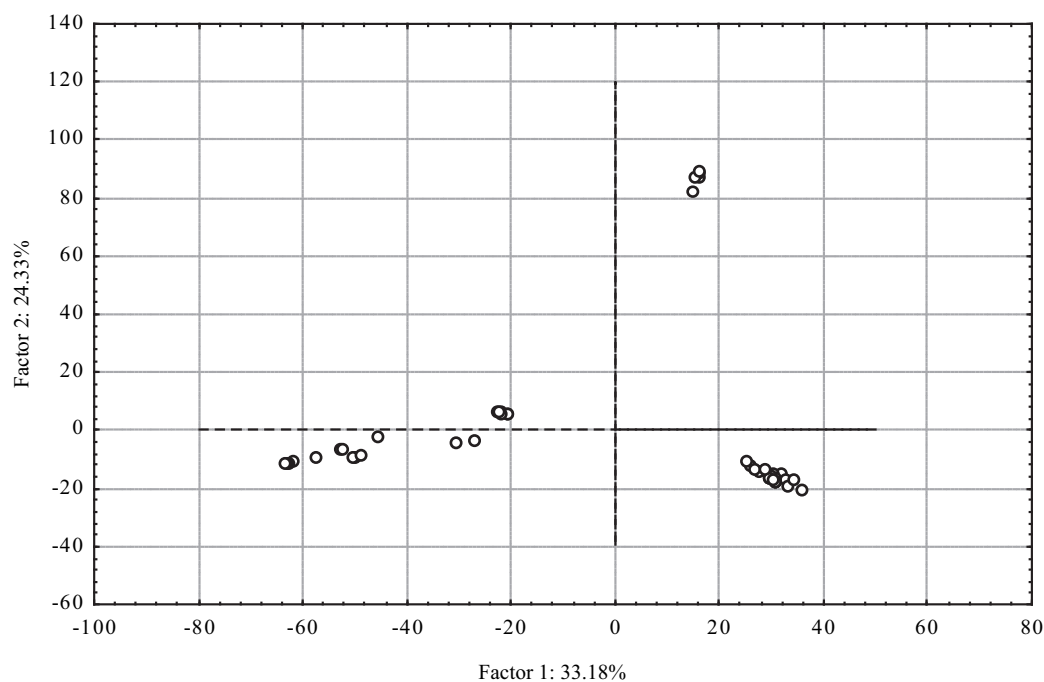


Figura 1— *Thymus zygis* subsp. *zygis*. Representación de los dos primeros componentes principales. Matriz de covarianza.

Se exponen a continuación para cada especie analizada los resultados del análisis de las muestras recogidas, el estudio exploratorio llevado a cabo mediante PCA y, para cada uno de sus compuestos (variables), los valores D_{mah} obtenidos como estimador de la separación de los dos grupos formados mediante el procedimiento de las k-medias, que se compararán con los que aparecen en la Tabla 1 para el correspondiente número de muestras.

Thymus zygis subsp. *zygis*

Datos de concentración: 48 muestras, que incluían hojas y flores tomadas de plantas individuales de *Thymus zygis* subsp. *zygis* de la Comunidad de Madrid se analizaron por DTD-GC-MS, dando lugar a una matriz de concentraciones relativas para 14 compuestos analizados. Una relación de estos compuestos y un resumen de los valores de sus concentraciones figuran en la Tabla 2. Los datos incluidos en la tabla indican la presencia de seis componentes (linalol, timol, carvacrol, geraniol y los acetatos de α -terpinilo y geraniol), que alcanzan concentraciones superiores al 70 % en algunas

muestras, pero que no fueron detectados en otras. Se trata por tanto de posibles marcadores de quimiotipos, sin descartar los restantes.

Análisis exploratorio multivariante: El análisis de componentes principales permite un estudio exploratorio multivariante de los datos considerados de forma conjunta. La contribución a la varianza de los tres primeros componentes principales es, respectivamente, del 33.2%, 24.3% y 23.0%. Los resultados de su aplicación a los datos de concentración de *Thymus zygis* subsp. *zygis* se muestran gráficamente en la Fig. 1 para los dos primeros componentes principales. Aunque se observa la presencia de grupos, la figura no puede tomarse como base para su definición. En primer lugar, no existe una clara definición del número de grupos presentes, pudiéndose señalar tres o cuatro de acuerdo con la figura, pero otros componentes no incluidos en la representación (del 3 en adelante) podrían poner de manifiesto grupos adicionales. Es difícil individualizar los compuestos incluidos en las agrupaciones, ya que los componentes empleados en la representación son combinación lineal de los compuestos (variables) de la matriz. En el caso del grupo de la zona superior de la figura, el compuesto predominante es claramente el linalol, varia-

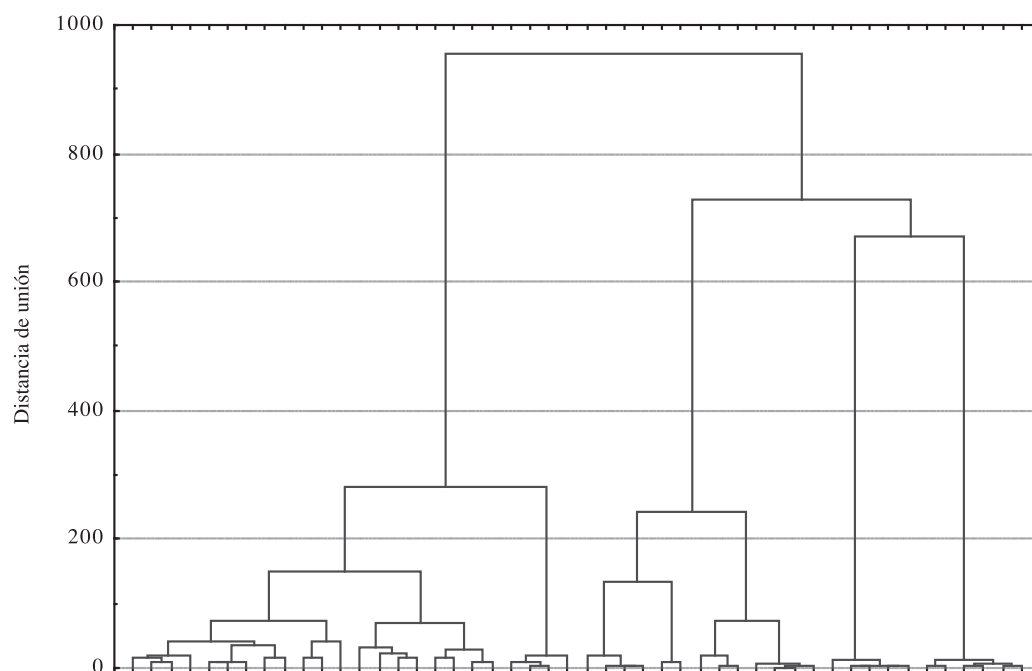


Figura 2— *Thymus zygis* subsp. *zygis*. Dendrograma jerárquico. Método de Ward, distancias euclídeas.

ble más importante del componente 2. Pero la separación a lo largo del primer componente depende de las concentraciones de p-cimeno, timol, carvacrol, acetato de α -terpinilo y geraniol, siendo difícil la asignación por grupos. Por último, los compuestos no correspondientes a ningún quimiotipo pueden actuar como ruido de fondo, aunque este problema es en esta serie de muestras menos importante por la gran variabilidad de algunos de los componentes. El diagrama jerárquico de la Fig. 2 es difícil de interpretar por el mismo motivo en términos de quimiotipos. Por otro lado, aunque se observan grupos claros, es imposible señalar a partir de qué distancia pueden ser significativos. Estos grupos marcados no siempre coinciden con los de la Fig. 1; por ejemplo, el grupo del centro-derecha de la Fig. 2 incluye las muestras representadas a la izquierda de la Fig. 1, pero también algunas de las del centro de dicha figura.

Estimación de la existencia de quimiotipos: De la Tabla 1 se obtiene el valor para 50 muestras (se han analizado 48) de la media (2.799) y desviación típica (0.282) de la distancia de Mahalanobis para una distribución normal. De acuerdo con lo mencionado en *Corrección de Bonferroni por número de compuestos*, para un nivel de confianza del 99.99% y 14 compuestos

$\alpha = 0.0001/14 = 0.00000714$ y el valor crítico es $2.799 + 0.282 \cdot 4.3395 = 4.02$. En la columna de la derecha de la Tabla 2, se señalan con un asterisco los compuestos con D_{mah} superiores a 4.02 y con dos asteriscos los superiores a 8.04. Se observa que la existencia de agrupaciones es significativa con excepción de los dos primeros compuestos, y muy clara para algunos de ellos (linalol, α -terpineol, acto. α -terpinilo). Un estudio de los datos individuales de estos dos últimos compuestos, estructuralmente relacionados, indica una coincidencia total entre los grupos de seis muestras formados.

Thymus zygis subsp. *sylvestris*

Datos de concentración: En la Tabla 3 se resumen los datos de concentración de 15 compuestos obtenidos mediante DTD-GC-MS en 50 muestras de *Thymus zygis* subsp. *sylvestris*, recogidas en la Comunidad de Madrid. Puede observarse, al igual que para el *Thymus zygis* subsp. *zygis*, una gran variabilidad. Siete compuestos presentan, al menos en una de las muestras, una concentración superior al 30% del total de volátiles, mientras que en otras presentan concentraciones muy bajas o

Tabla 2

Compuestos determinados por DTD-GC-MS en 48 muestras de *Thymus zygis* subsp. *zygis*. Valores mínimo, medio y máximo de sus concentraciones porcentuales. Grupos (media y número de muestras) obtenidos mediante el procedimiento de las k-medias, distancia de Mahalanobis D_{mah} entre grupos y significación estadística de los grupos hallados. * valor mayor que el valor crítico para el nivel de significación $\alpha = 0.0001$, ** valor mayor que el doble del valor crítico para el nivel de significación $\alpha = 0.0001$.

<i>Thymus zygis</i> subsp. <i>zygis</i>								
	V. Medio	V. Mínimo	V. Máximo	Media G1	NG1	Media G2	NG2	D_{mah}
Mirceno	1.1	0	4.0	0.4	20	1.61	28	2.46
p-Cimeno	12.3	0	47.1	2.4	29	27.45	19	3.96
γ -Terpineno	4.9	0	22.3	1.5	35	14.0	13	4.12 *
Linalol	11.3	0	98.0	1.5	43	95.1	5	55.20 **
Timol	13.5	0	70.4	4.1	38	49.0	10	4.67 *
Carvacrol	15.7	0	74.9	1.9	31	41.0	17	4.58 *
β -Pino	0.8	0	6.3	0.6	46	5.1	2	5.08 *
Alcanfor	0.4	0	3.5	0.3	46	3.2	2	8.05 **
Terpinen-4-ol	1.6	0	9.6	0.6	39	5.8	9	5.30 *
α -Terpineol	1.3	0	12.6	0.1	42	9.8	6	12.22 **
Acto. α -terpinilo	12.9	0	87.3	2.8	42	84.0	6	11.38 **
Geraniol	17.4	0	86.8	1.4	37	71.5	11	9.54 **
Acto. geraniol	5.8	0	76.3	2.8	46	73.4	2	7.62 *
1,8-cineol	0.6	0	4.3	0.3	41	2.4	7	5.13 *

Tabla 3

Compuestos determinados por DTD-GC-MS en 50 muestras de *Thymus zygis* subsp. *sylvestris*. Valores mínimo, medio y máximo de sus concentraciones porcentuales. Grupos (media y número de muestras) obtenidos mediante el procedimiento de las k-medias, distancia de Mahalanobis D_{mah} entre grupos y significación estadística de los grupos hallados. * valor mayor que el valor crítico para el nivel de significación $\alpha = 0.0001$, ** valor mayor que el doble del valor crítico para el nivel de significación $\alpha = 0.0001$.

<i>Thymus zygis</i> subsp. <i>sylvestris</i>								
	V. Medio	V. Mínimo	V. Máximo	Media G1	NG1	Media G2	NG2	D_{mah}
Mirceno	4.0	0.7	34.4	3.0	48	28.1	2	6.66 *
p-Cimeno	22.1	1.8	47.3	15.5	30	32.0	20	2.89
γ -Terpineno	10.3	0.3	32.7	7.8	37	17.6	13	2.64
Linalol	3.8	0.0	75.1	—	—	—	—	—
Timol	41.7	0.0	69.8	14.8	12	50.2	38	3.4
Carvacrol	1.8	0.0	10.0	1	35	3.7	15	2.36
β -Pino	2.1	0.2	12.7	1.6	45	6.6	5	3.64
Alcanfor	2.7	0.0	18.4	1.9	43	7.4	7	2.62
Terpinen-4-ol	1.8	0.0	9.4	1.1	44	6.7	6	5.88 *
α -Terpineol	0.7	0.0	30.1	—	—	—	—	—
Acto. α -terpinilo	1.9	0.0	51.2	0	48	46.9	2	53.43 **
1,8-Cineol	1.9	0.3	17.8	1.2	48	17.5	2	12.34 **
α -Pino	1.6	0.1	15.2	-	-	-	-	-
Canfeno	2.1	0.2	12.7	1.7	45	6.6	5	3.61
Cariofileno	1.3	0.2	3.1	0.9	30	1.8	20	2.49

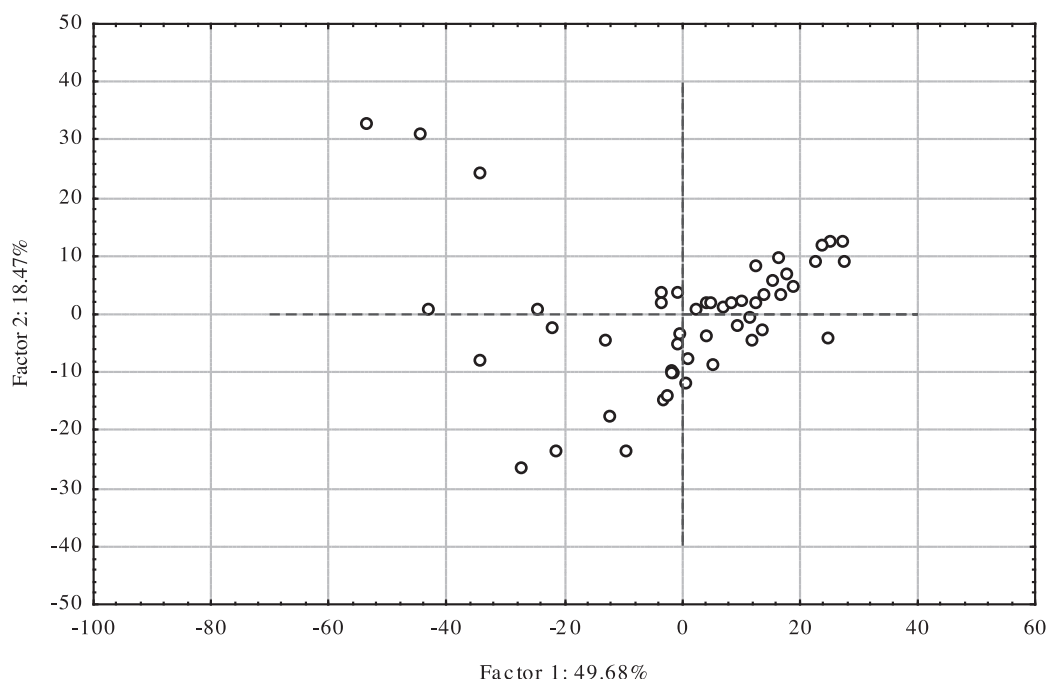


Figura 3— *Thymus zygis* subs. *sylvestris*. Representación de los dos primeros componentes principales. Matriz de covarianza.

nulas. El porcentaje de la varianza explicado por los tres primeros componentes principales es, respectivamente, del 33.2%, el 24.3% y el 23.0%. La representación gráfica que utiliza los dos primeros (Fig. 3) no muestra signos claros de agrupación, ya que aunque las tres muestras de la parte superior izquierda de la figura están claramente separadas, sus componentes mayoritarios son distintos (terpineol, acetato de terpinilo y linalol), tratándose por tanto de muestras aisladas en cuanto a su composición. Los resultados del análisis cluster son, al igual que en el caso del *T. zygis* subsp. *zygis*, de poca utilidad al incluir información conjunta sobre todos los datos.

Estimación de la existencia de quimiotipos: De la Tabla 1 se obtiene como en el caso anterior el valor para 50 muestras de la media (2.799) y desviación típica (0.282) de la distancia de Mahalanobis para una distribución normal. Para un nivel de confianza del 99.99% y 15 compuestos $\alpha = 0.0001/15 = 0.00000666$ y el valor crítico es $2.799 + 0.282 \cdot 4.3544 = 4.03$. En la columna de la derecha de la Tabla 3, se señalan con un asterisco los compuestos con D_{mah} superiores a 4.03 y con dos asteriscos los superiores a 8.06. En este caso, sólo cuatro compuestos presentan agrupaciones significativas.

Los dos casos para los que D_{mah} tiene un mayor valor (acetato de α -terpinilo y 1,8-cineol), lo mismo que el mircenol, corresponden a grupos formados por tan sólo dos muestras: para el terpinen-4-ol, un posible quimiotipo incluiría seis muestras, pero con un valor medio de concentración no muy elevado (6.7%). Tres compuestos, linalol, α -terpineol y α -pineno no han podido incluirse al obtenerse mediante el procedimiento de las k-medias un grupo de una sola muestra, para el que no se puede calcular el valor de D_{mah} .

Lavandula luisieri

Datos de concentración: En el caso de *Lavandula luisieri*, las muestras de hojas y flores de cada uno de los 50 individuos se analizaron por separado, determinándose en cada una de ellas la concentración de 9 compuestos. Un resumen de los resultados, con los valores medios, mínimo y máximo de concentración para cada compuesto, figura en las Tablas 4 (flores) y 5 (hojas). El compuesto C1 se identificó como 5-metilen-2,3,4,4-tetrametilclopent-2-enone (Baldovini *et al.* 2005), mientras que C3 parece ser un hidroxido derivado de la cetona C1 y C2 el

acetato del compuesto C3, y N1 y N2 son respectivamente acetatos de *trans*- y *cis*-a-necrodilo. Estos compuestos, junto con C1, C2 y C3, son característicos de la *Lavandula luisieri*. La determinación de estos compuestos y los resultados obtenidos se detallan en Sanz *et al.* (2004) y en González-Coloma *et al.* (2006). Los valores de la Tablas 4 y 5 indican una composición similar para hojas y flores, por lo que los resultados se exponen de forma conjunta. Al igual que en las especies de *Thymus*

analizadas, la diversidad en la concentración de los componentes es muy marcada, estando ausentes en algunas muestras los compuestos mayoritarios de otras.

El análisis de componentes principales da lugar también a resultados similares para flores y hojas. Mientras que para las hojas los tres primeros componentes principales explican, respectivamente, el 64.1%, 28.5% y 4.2% de la varianza, para los volátiles de las flores estos porcentajes son del 64.4%, 21.8% y 8.9%. Las representaciones de

Tabla 4

Compuestos determinados por DTD-GC-MS en 50 muestras de flores de *Lavandula luisieri*. Valores mínimo, medio y máximo de sus concentraciones porcentuales. Grupos (media y número de muestras) obtenidos mediante el procedimiento de las k-medias, distancia de Mahalanobis D_{mah} entre grupos y significación estadística de los grupos hallados. * valor mayor que el valor crítico para el nivel de significación $\alpha = 0.0001$, ** valor mayor que el doble del valor crítico para el nivel de significación $\alpha = 0.0001$

<i>Lavandula luisieri</i> (Flores)								
	V. Medio	V. Mínimo	V. Máximo	NG1	Media G1	NG2	Media G2	D_{mah}
Fenchona	6.7	0.0	64.2	44	2.5	6	36.8	5.13 *
Alcanfor	28.5	0.0	87.8	34	7.6	16	72.8	6.48 *
1,8-cineol	19.9	0.0	85.2	33	3.5	17	51.7	4.55 *
C1	16.5	0.4	60.0	38	10.7	12	34.9	3.29
C2	17.2	0.9	52.3	41	12.1	9	40.4	4.1 *
C3	4.0	0.5	13.4	42	2.9	8	9.8	3.84
N1	0.2	0.0	4.2	48	0.1	2	3.0	10.89 **
N2	0.3	0.0	1.2	37	0.2	13	0.8	5 *
Hidroxycadinena	6.7	0.0	38.1	46	5.1	4	25.7	4.31 *

Tabla 5

Compuestos determinados por DTD-GC-MS en 50 muestras de hojas de *Lavandula luisieri*. Valores mínimo, medio y máximo de sus concentraciones porcentuales. Grupos (media y número de muestras) obtenidos mediante el procedimiento de las k-medias, distancia de Mahalanobis D_{mah} entre grupos y significación estadística de los grupos hallados. * valor mayor que el valor crítico para el nivel de significación $\alpha = 0.0001$, ** valor mayor que el doble del valor crítico para el nivel de significación $\alpha = 0.0001$.

<i>Lavandula luisieri</i> (Hojas)								
	V. Medio	V. Mínimo	V. Máximo	Media G1	NG1	Media G2	NG2	D_{mah}
Fenchona	2.9	0.0	28.4	2.1	48	23.3	2	5.77 *
Alcanfor	17.1	0.0	80.9	1.6	36	57.0	14	6.89 *
1,8-cineol	23.3	0.0	76.7	5.7	27	44.0	23	3.57
C1	24.1	4.6	60.9	18.9	38	40.6	12	3.29
C2	22.0	0.3	52.6	17	38	37.7	12	3.03
C3	8.3	1.5	16.8	5.4	30	12.6	20	3.19
N1	0.4	0.0	2.9	0.2	46	2.1	4	6.39 *
N2	0.6	0.0	2.0	0.3	34	1.2	16	3.28
Hidroxycadinena	1.3	0.0	6.2	0.7	41	4.3	9	4.26 *

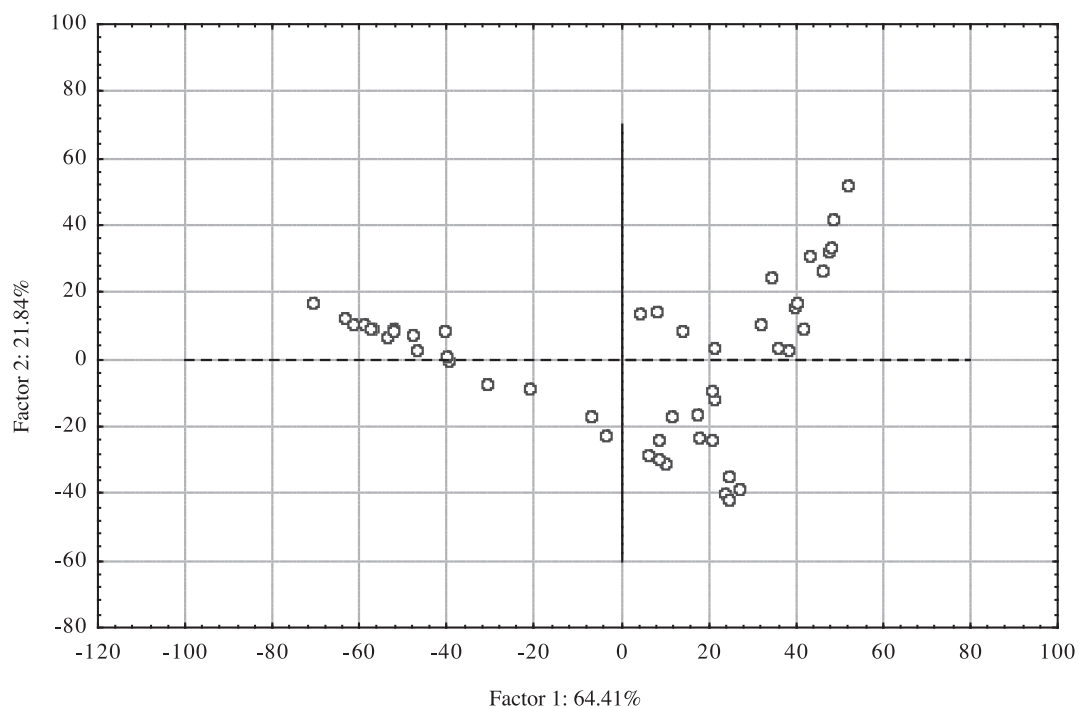


Figura 4— *Lavandula luisieri* (flores). Representación de los dos primeros componentes principales. Matriz de covarianza.

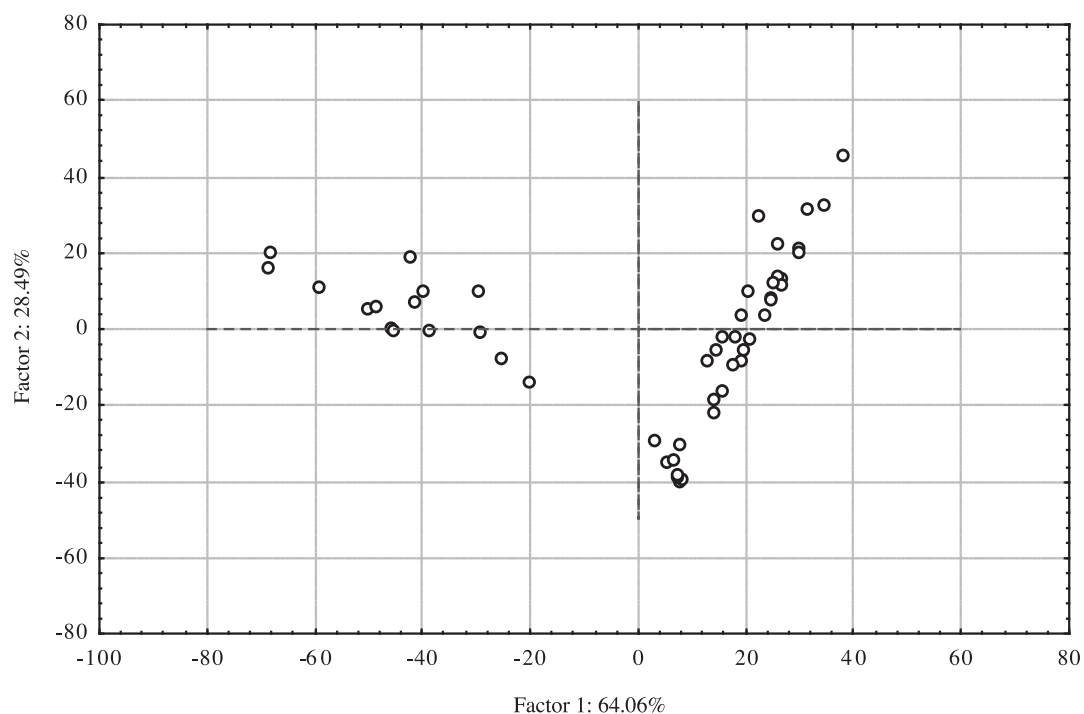


Figura 5— *Lavandula luisieri* (hojas). Representación de los dos primeros componentes principales. Matriz de covarianza.

componentes principales de las Figuras 4 y 5 indican que las muestras se distribuyen de forma estructurada. Sin embargo, y a diferencia de lo observado para *Thymus zygis* subsp. *zygis*, para los compuestos más importantes (alcanfor, 1,8-cineol) la distribución no muestra discontinuidades (Sanz *et al.* 2004). Por ejemplo, la concentración de alcanfor y de los compuestos C aumenta para las muestras representadas en ambas figuras de izquierda a derecha, mientras que la de 1,8-cineol disminuye, y las muestras de la parte superior de la figura presentan altas concentraciones de 1,8-cineol y/o alcanfor, que disminuyen de forma continua hacia la parte inferior de la tabla.

Estimación de la existencia de quimiotipos para muestras de flores y de hojas de *Lavandula luisieri*. De la Tabla 1 se obtiene como en el caso anterior el valor para 50 muestras de la media (2.799) y desviación típica (0.282) de la distancia de Mahalanobis para una distribución normal. Para un nivel de significado del 99.99% y 9 compuestos $\alpha' = 0.0001/9 = 0.00001111$ y el valor crítico es $2.799 + 0.282 \cdot 4.2411 = 3.99$. En la columna de la derecha de las Tablas 4 y 5, se señalan con un asterisco los compuestos con D_{mah} superiores a 3.99 y con dos asteriscos los superiores a 7.98. Las tendencias que se observan en las Figuras 4 y 5 no se ponen de manifiesto en una clara definición de quimiotipos. El único caso en que se obtiene un valor de D_{mah} superior al doble del crítico es para el compuesto N2, con un grupo formado por dos muestras cuyo valor medio de concentración no supera el 3%. En el caso de las hojas, compuestos como el 1,8-cineol (que varía entre 0 y 76%), el C1 (entre 4.5 y 60.9%) y el C2 (entre un 0.3 y un 52.6%), no alcanzan el nivel de significación marcado.

CONCLUSIONES

El método descrito permite, utilizando siempre criterios objetivos, el distinguir a partir de datos de composición química la posible presencia de quimiotipos y el asignarles un nivel cuantitativo de significado. La elección del procedimiento de las k-medias como técnica de agrupamiento parece indicada, ya que el número de grupos debe ser siempre igual a dos para este objetivo. La asignación de significado a esa agrupación requiere el uso de un parámetro estadístico que tenga en cuenta tanto la separación entre grupos como su dispersión, para lo que la distancia de Mahalanobis presenta ventajas, aunque podrían usarse otros parámetros. El parámetro de Hopkins (Hopkins 1954), que

se ha propuesto como una medida previa de la tendencia a la formación de agrupaciones, es útil para descartar la presencia de distribuciones uniformes, pero en nuestro caso no parece aplicable, ya que proporciona resultados positivos en cuanto a agrupación tanto para distribuciones unimodales como para las bimodales. La aplicación del método a datos de composición obtenidos en nuestro laboratorio por DTD-GC-MS para *Thymus zygis* subsp. *zygis*, *Thymus zygis* subsp. *sylvestris* y *Lavandula luisieri* ha permitido estudiar su comportamiento para tres diferentes casos de distribución de componentes volátiles en plantas.

En el caso del *Thymus zygis* subsp. *zygis*, en el que la distribución indica una separación en grupos muy clara, el método confirma la interpretación subjetiva de los datos, pero también asigna valores cuantitativos y niveles de significado a los distintos compuestos característicos, que pueden utilizarse con fines comparativos. Para el *Thymus zygis* subsp. *sylvestris*, en el que la presencia de quimiotipos es menos clara, el método también proporciona información cuantitativa sobre ellos que puede utilizarse en su valoración. El significado de las muestras individuales con composición diferente a la del resto que aparecen en este caso requiere un tratamiento estadístico diferente, que podría estar basado en los métodos de detección de valores atípicos (detección de outliers). Sin embargo, el que esta diferencia en la composición fuese estadísticamente significativa no implicaría la existencia de quimiotipos definidos como “grupos de individuos de una especie”, sino sólo el interés de ampliar el estudio analizando un número superior de muestras.

Los datos obtenidos para la composición de la *Lavandula luisieri* presentan un diferente tipo de distribución. La concentración de algunos compuestos parece no seguir tendencias unimodales o bimodales, sino que sus valores aparecen con una frecuencia constante a lo largo del intervalo, siguiendo una distribución uniforme. La aplicación del método, usando k-medias para la definición de grupos y D_{mah} para estudiar su diferencia, indica la presencia significativa de algunos quimiotipos: la distribución uniforme podría señalar la existencia de una hibridación limitada entre ellos.

En las tres especies mencionadas se presentan por otro lado correlaciones entre concentraciones de algunos componentes. El método estadístico propuesto requeriría una redefinición del método de agrupación que, manteniendo sus ventajas, incluyera la consideración de la magnitud relativa de las concentraciones de los compuestos, de sus correlaciones sobre todo para casos de origen biosintético

común y, sobre todo, de las tendencias que pueden presentarse de acuerdo con la distribución de los compuestos en plantas de una misma especie.

AGRADECIMIENTOS

Este trabajo ha sido subvencionado en parte por la Comunidad de Madrid (Programa ANALISYC, S505/AGR-0312).

BIBLIOGRAFÍA

- BALDOVINI, N.; LAVOINE-HANNEGUELLE, S.; FERRANDO, G.; DUSART, G. & LIZZANI-CUVELIER, L. 2005. Necrodane monoterpenoids from *Lavandula luisieri*. *Phytochemistry* 66: 1651-1655.
- ESTEBAN, J. L.; MARTÍNEZ-CASTRO, I. & SANZ, J. 1993. Evaluation and optimization of the automatic thermal desorption (ATD) method in the gas chromatographic determination of plant volatile compounds. *J. Chromatogr. A* 657: 155-164.
- GARCÍA-MARTÍN, D. & GARCÍA-VALLEJO, M. C. 1979. Las esencias de la Farmacopea española 9ª edición: III. Esencia de Tomillo. Segundas Jornadas Farmacéuticas, Málaga, 459.
- GARCÍA-MARTÍN, D. & GARCÍA-VALLEJO, M. C. 1983. Chemotypes of *Thymus zygis* (Lofl.) L. of Guadarrama Sierra and other places in Castille (Spain). IX International Congress of Essential Oils, 13-17 Marzo, Singapur.
- GARCÍA-VALLEJO, M. C.; REBOLLAR-REIER, M. P. & GARCÍA-MARTÍN, D. 1992. Composición química del aceite esencial de *Thymus vulgaris* L. en la Comunidad de Madrid. Ministerio de Agricultura, Pesca y Alimentación, INIA, Madrid, 221.
- GARCÍA-VALLEJO, M. I. 1992. Aceites esenciales de las lavándulas ibéricas: ensayo de la quimiotaxonomía. Tesis Doctoral, Universidad Complutense de Madrid, Facultad de Biología.
- GONZÁLEZ-COLOMA, A.; MARTÍN-BENITO, D.; MOHAMED, N.; GARCÍA-VALLEJO, M. C. & SORIA, A.C. 2006. Antifeedant effects and chemical composition of essential oils from different populations of *Lavandula luisieri* L. *Biochem. Syst. Ecol.* 34: 609-616.
- HARTIGAN, J. A. 1975. Clustering Algorithms. Wiley. New York.
- HARTIGAN, J. A. & WONG, M. A. 1979. A K-Means Clustering Algorithm. *Appl. Stat.* 28: 100-108.
- HOPKINS, B. 1954. A new method for determining the type of distribution of plant individuals. *Ann. Bot.* 18: 213-227.
- LAWSON, R. G. & JURIS, P.C. 1990. New index for clustering tendency and its application to chemical problems. *J. Chem. Inf. Comp. Sci.* 30: 36-41.
- MACQUEEN, J. B. 1967. Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1, 281-297.
- MAHALANOBIS, P. C. 1936. On the generalized distance in statistics. Proceedings of the National Institute of Science of India, 12, 49-55.
- MCLAFFERTY, F. W. & STAUFFE, D. B. 1989. Wiley/NBS Registry of Mass Spectral Data, Wiley, New York.
- MICHET, A.; CHALCHAT, J.; FIGUÉREDO, G.; THÉBAUD, G.; BILLY, F. & PÉTEL, G. 2008. Chemotypes in the volatiles of wild thyme (*Thymus pulegioides* L.). *J. Essent. Oil Res.* 20: 101-103.
- MORALES VALVERDE, R. 1986. Taxonomía del género *Thymus* (excluida la sección *Serpyllum*) y *Timbra* en la Península Ibérica. *Ruizia* 3: 1-324.
- NIST/EPA/NIH. 2002. Mass spectral library with search program, data version: NIST'02, Software Version 2.0.
- SÁEZ, F. 1995. Essential oil variability of *Thymus zygis* growing wild in Southeastern Spain. *Phytochemistry* 40: 819-825.
- SALGUEIRO, L. R. 1994. Os tomillos portugueses e os seus óleos essenciais. Tesis Doctoral, Universidad de Coimbra, Facultad de Farmacia, Coimbra.
- SANZ, J.; SORIA, A. C. & GARCÍA-VALLEJO, M. C. 2004. Analysis of volatile components of *Lavandula luisieri* L. by direct thermal desorption-gas chromatography-mass spectrometry. *J. Chromatogr. A* 1024: 139-146.
- SHAFER, J. P. 1995. Multiple hypothesis testing. *Ann. Rev. Psych.* 46: 561-584.
- STAHL-BISKUP, E. 1991. The chemical composition of *Thymus* oils: a review of the literature 1960-1989. *J. Ess. Oil Res.* 3: 61-82.
- STATISTICA (data analysis software system, 2005). StatSoft, Inc. version 7.1. www.statsoft.com.
- VELASCO NEGUERUELA, A. & PÉREZ ALONSO, M. J. 1984. Aceites esenciales. *Anal. Bromatol.* 36: 301-308.
- VELASCO NEGUERUELA, A. & PÉREZ ALONSO, M. J. 1985. Aceites esenciales de tomillos ibéricos. II. Contribución al conocimiento del aceite esencial de *Thymus lacaitae* Pau. *Anal. Jard Bot. Madrid*, 42, 159.